EXHIBIT B

Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening

PEIZHI LUO, ROBERT J. HAYES, CHERYL CHAN, DIANE M. STARK, MARIAN Y. HWANG, JONATHAN M. JACINTO, PADMAJA JUVVADI, HELEN S. CHUNG, ANIRBAN KUNDU, MARIE L. ARY, AND BASSIL I. DAHIYAT
Xencor, Inc., Monrovia, California 91016, USA

(RECEIVED November 13, 2001; FINAL REVISION February 11, 2002; ACCEPTED February 13, 2002)

Abstract

Granulocyte-colony stimulating factor (G-CSF) is used worldwide to prevent neutropenia caused by high-dose chemotherapy. It has limited stability, strict formulation and storage requirements, and because of poor oral absorption must be administered by injection (typically daily). Thus, there is significant interest in developing analogs with improved pharmacological properties. We used our ultrahigh throughput computational screening method to improve the physicochemical characteristics of G-CSF. Improving these properties can make a molecule more robust, enhance its shelf life, or make it more amenable to alternate delivery systems and formulations. It can also affect clinically important features such as pharmacokinetics. Residues in the buried core were selected for optimization to minimize changes to the surface, thereby maintaining the active site and limiting the designed protein's potential for antigenicity. Using a structure that was homology modeled from bovine G-CSF, core designs of 25–34 residues were completed, corresponding to 10^{21}–10^{28} sequences screened. The optimal sequence from each design was selected for biophysical characterization and experimental testing; each had 10–14 mutations. The designed proteins showed enhanced thermal stabilities of up to 13°C, displayed five- to 10-fold improvements in shelf life, and were biologically active in cell proliferation assays and in a neutropenic mouse model. Pharmacokinetic studies in monkeys showed that subcutaneous injection of the designed analogs results in greater systemic exposure, probably attributable to improved absorption from the subcutaneous compartment. These results show that our computational method can be used to develop improved pharmaceuticals and illustrate its utility as a powerful protein design tool.

Keywords: Protein design; computational screen; stability; cytokines; granulocyte-colony stimulating factor

Many techniques have been used in the design of new and improved proteins. In vitro directed evolution methods such as phage display, DNA shuffling, and error-prone PCR are widely used. Rational design approaches continue to be applied, and strategies that combine both are now being used. Successful designs include enzymes (Chen and Arnold 1991; Stemmer 1994; Zhao et al. 1998) and other proteins (Crameri et al. 1996), as well as therapeutically useful proteins such as hormones and cytokines (Lowman and Wells 1993; Heikoop et al. 1997; Grossmann et al. 1998; Chang et al. 1999). The experimental techniques involve the generation and screening of libraries of random protein sequences. However, the number of sequences that can be screened experimentally is limited (about 10^{14} for library panning and 10^{7} for high throughput screening). Libraries of this size allow for the simultaneous modification of only about 10 residues.

Reprint requests to: Bassil I. Dahiyat, Xencor, Inc., 111 W. Lemon Avenue, Monrovia, California 91016, USA; e-mail: baz@xencor.com; fax: (626) 256-3562.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.4580102.

Computational methods have also been used that perform in silico screening of protein sequences (Hellinga and Richards 1994; Desjarlais and Handel 1995; Dahiyat and Mayo 1996, 1997a; Street and Mayo 1999; Jiang et al. 2000; Kraemer-Pecore et al. 2001; Pokala and Handel 2001). Exploiting the efficiency and speed of computers, these methods can randomly screen a vast number of sequences (up to 10^{80}), allowing for the simultaneous consideration and modification of more than 60 residues. Searching such large sequence spaces drastically improves the possibility of finding novel protein sequences with improved properties.

Investigators have recently developed a computational screening method that finds the optimal sequence for a defined three-dimensional structure, allowing all or part of the sequence to change (Dahiyat and Mayo 1996). This method, termed Protein Design Automation (PDA), scores the fit of sequences to the three-dimensional structure using physical-chemical potential functions that model the energetic interactions of protein atoms, including steric, solvation, and electrostatic interactions. PDA couples these potential functions with a highly efficient search algorithm to accurately screen up to 10^{80} sequences. Because the screening is performed in silico, multiple simultaneous mutations can be made, and novel sequences that are very different from wild type can be discovered. The method has been validated by numerous experimental tests and has resulted in the design of new proteins with improved stability and conformational specificity, and novel activity (Dahiyat and Mayo 1996, 1997a; Malakauskas and Mayo 1998; Strop and Mayo 1999; Shimaoka et al. 2000; Bolon and Mayo 2001; Marshall and Mayo 2001).

PDA also has the advantage of being able to control the location and type of mutations. For example, the design can be limited to the hydrophobic core. Mutations in the core can produce significant improvements in protein stability but do not change binding epitopes on the surface of the molecule. Thus, the molecular surface can be kept identical to the native structure, retaining biological activity and limiting toxicity and antigenicity. This feature is particularly important in the design of therapeutic proteins.

We wanted to take advantage of these features of PDA and explore its utility in the design of improved pharmaceuticals. We therefore used PDA as an ultrahigh throughput screen for improved analogs of a therapeutic protein, granulocyte-colony stimulating factor (G-CSF). G-CSF is a hematopoietic growth factor of 174 residues that induces differentiation and proliferation of granulocyte-committed progenitor cells. It is used clinically to treat cancer patients and alleviate the neutropenia induced by high-dose chemotherapy. G-CSF belongs to the class of long-chain four-helix bundle cytokines that bind asymmetrically to homodimeric complexes of cell-surface receptors to initiate an intracellular signaling cascade. Their structural similarity allows the design strategy chosen for G-CSF to be immediately applicable to the other four-helix bundle cytokines (human growth hormone, erythropoietin, the interleukins, and interferon-α/β—all clinically important compounds) and thus broadens the potential impact of the results.

Although the cytokines are functionally very efficacious, their pharmacological properties are not ideal. For example, G-CSF, like most proteins, is not absorbed orally to any significant extent and must be administered by frequent (daily) injections throughout the course of treatment. It also has limited stability and strict formulation and storage requirements, including the need to be kept refrigerated. Thus, there is significant interest in developing analogs with improved pharmacological properties.

We sought to use PDA to improve the physicochemical characteristics of G-CSF. Improving these properties can make a molecule more robust, enhance its shelf life, or make it more amenable to use in alternate delivery systems and formulations. It can also affect clinically important features such as pharmacokinetics and result in a drug that is safer for human use. Our design strategy was to optimize the core to improve the stability and solution properties of G-CSF while preserving receptor binding and biological activity.

The template structure used for in silico screening was a homology model of human G-CSF in which the human sequence was mapped onto bovine G-CSF. We designed several novel core sequences, cloned and expressed them, characterized their stabilities, tested them for functional activity both in vitro and in vivo, and studied their pharmacokinetics in monkeys. The designed proteins showed enhanced thermal stabilities, displayed five- to 10-fold improvements in shelf life, and were biologically active both in cell proliferation assays and in a neutropenic mouse model. Subcutaneous injection of the most stable variant in monkeys also resulted in greater systemic exposure, probably attributable to improved absorption from the subcutaneous compartment. These results indicate that PDA has great potential as a powerful in silico tool in the design of improved pharmaceutical proteins.

Results and Discussion

Homology modeling

The crystal structure of bovine G-CSF (PDB record 1bgc) (Lovejoy et al. 1993) was used as the starting point for modeling because the crystal structure of human G-CSF (PDB record 1rhg) (Hill et al. 1993) is at a lower resolution and is missing key fragments, including a structurally important disulfide bond between positions 64 and 74. Bovine G-CSF is a good model for human G-CSF because the sequences are the same length and 142 of 174 amino acids are identical (82%). The residues that differ in the bovine sequence were replaced with the human residues for those

positions, and the conformations of the replaced side chains were optimized using PDA. Most of the replaced residues were solvent exposed, thereby introducing little strain into the structure and allowing typical PDA parameters to be used for conformation optimization. One substitution, however, was at a buried site, G167V, and clashed sterically with a nearby disulfide bond. To accommodate the larger Val, the side-chain conformation at this position was optimized using a less restrictive van der Waals scale factor (0.6 instead of 0.9). The entire structure was then briefly minimized to relax the strain. The final structure that served as the template for all the designs is shown in Figure 1.

Core designs

Unlike many experimental sequence screening methods, PDA allows control over which residues are allowed to change. Core residues were selected because optimization of these positions can improve stability yet minimize changes to the molecular surface, thus limiting the designed protein's potential for antigenicity. Ala scanning studies of G-CSF indicate one or two binding sites on the protein surface that are probably responsible for granulopoietic activity (Reidhaar-Olson et al. 1996; Young et al. 1997) (Fig. 1). Although recent crystallographic studies of G-CSF complexed to its receptor show only one binding site in a novel 2:2 complex (Horan et al. 1996; Aritomi et al. 1999), both sites were avoided in the core designs to ensure preservation of function.

Two PDA design calculations were run: a deep core design that included residues deeply buried in the interior of the protein and an expanded core design (exp_core) that also included less buried peripheral core residues. The deep core design had 26 core positions that were allowed to vary (shown yellow and gold in Fig. 2), whereas exp_core had 34 (shown yellow and turquoise in Fig. 2). Only hydrophobic amino acids were considered at the variable core positions. These included Ala, Val, Ile, Leu, Phe, Tyr, and Trp. Gly was also allowed for the variable positions that had Gly in the bovine wild-type structure (positions 28, 149, 150, and 167). Met and Pro were not allowed.

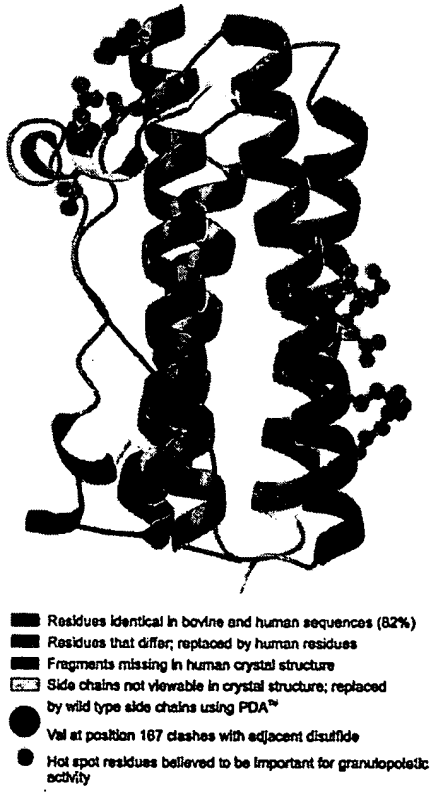

Fig. 1. Template structure of hG-CSF used for Protein Design Automation (PDA) designs. The human sequence was homology modeled onto the bovine crystal structure (PDB record 1bgc). The residues that differ in the bovine sequence or were not present in the bovine crystal structure were replaced with the residues from the human sequence. The conformations of the replaced side chains were optimized using PDA (the larger Val at position 167 was optimized using a less restrictive van der Waals scale factor), and the entire structure was energy minimized for 50 steps.

Optimal sequences

The optimal sequences selected by PDA are also shown in Figure 2. The optimal sequence from the deep core design had 10 mutations (named core10), and the optimal exp_core sequence had 11 (named exp_core11); thus, 33%–38% of the variable residues changed their identities. Eight of the mutated positions changed to the same amino acid in both designs. Changing the set of design positions can significantly impact the amino acid selected at a given position. For example, in the deep core design, Leu89 retains the same amino-acid identity and conformation as wild type. However, in the exp_core design, when Leu92 is also allowed to vary, both positions (Leu89 and Leu92) mutate to Phe, indicating a coupling between these two core residues. The modeled structure of the sequence selected in the deep core design (core10) is shown in Figure 3.

Native human G-CSF (met hG-CSF) and the optimal sequence from each of the core designs were cloned, expressed in *Escherichia coli*, and purified for experimental studies.

Thermal stability

The far-ultraviolet (UV) circular dichroism (CD) spectra for met hG-CSF and the designed proteins were nearly identical to each other and to published spectra for met hG-CSF (Reidhaar-Olson et al. 1996; Young et al. 1997), indicating highly similar secondary structure and tertiary folds (data

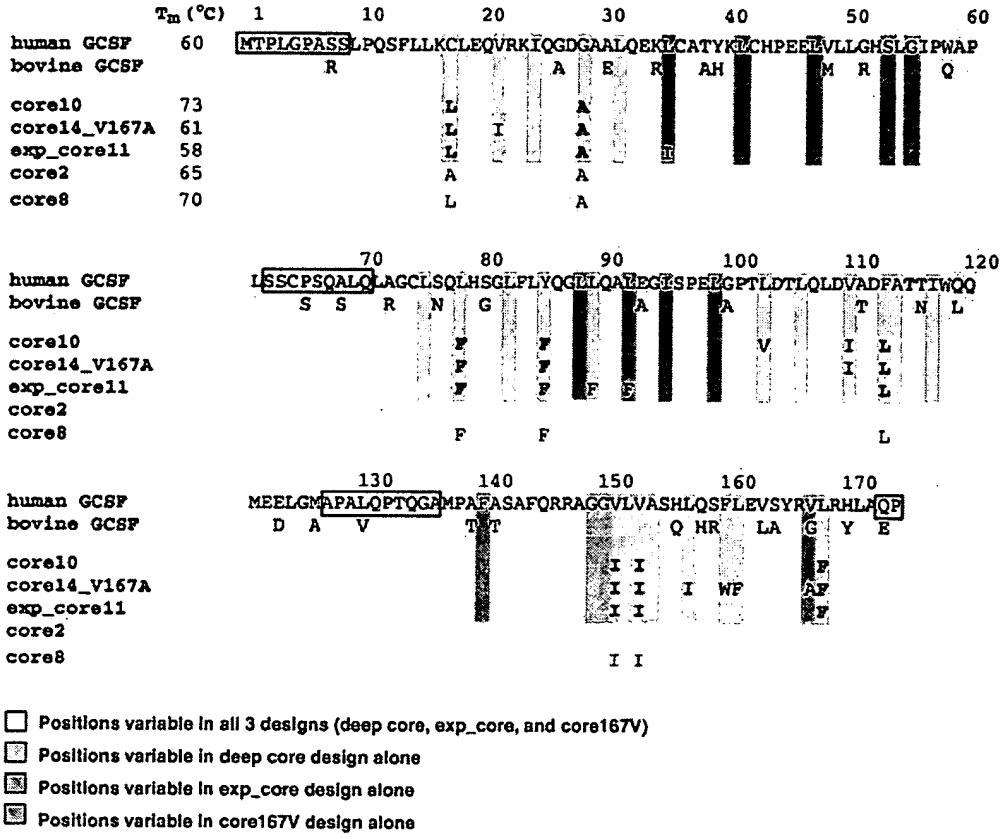

Fig. 2. Sequences of hG-CSF analogs. Native human and bovine sequences are shown at the *top*. The fragments missing in the crystal structure of the human sequence are shown boxed. Variable positions are colored. The deep core design had 26 variable positions. exp_core had 34, and core167V had 25. The optimal sequence from each design is shown. Letters indicate core residues that mutated relative to native hG-CSF; blanks indicate no change. Positions that changed to the same amino acid in all three core designs are indicated in bold. Core2 and core8 sequences were not obtained from PDA calculations but were derived by reverting some of the core10 mutations to wild type. Melting temperatures (T_ms) obtained for the designed proteins are also shown.

not shown). Thermal denaturation was monitored at 222 nm, and the melting temperatures (T_ms) were derived from the derivative curve of the ellipticity at 222 nm versus temperature (Fig. 4). Thermal denaturation of G-CSF and its variants is irreversible; however, T_m can be used to quickly assess the relative stability of different mutants. Stability under storage conditions, which is more relevant clinically, was evaluated with shelf-life studies (see below).

The T_m for met hG-CSF was 60°C, identical to that reported in other studies (Kolvenbach et al. 1997). Core10 showed an increase in stability of 13°C, whereas the T_m of exp_core11 was very similar to wild type (Fig. 2 and Fig. 4). The increased stability seen with core10 may be attributable to improved packing interactions and optimized hydrophobic burial of side chains. Other possibilities include decreased aggregation resulting from elimination of the free cysteine at position 17. The Gly to Ala mutation at position 28 caused a significant improvement in helical propensity that could also be the source of the improved stability.

Identifying critical mutations using derived sequences

To differentiate between these possibilities, two additional sequences derived from the core10 mutant sequence were made and their T_ms measured. One of these (core8) was identical to core10 except that two mutations distant from the others were reverted to wild type (L103V and V110I). These were the two positions that did not mutate in exp_core11. The T_m of core8 was 70°C, similar to core10, indicating that the mutations at 103 and 110 were not responsible for core10's improved stability.

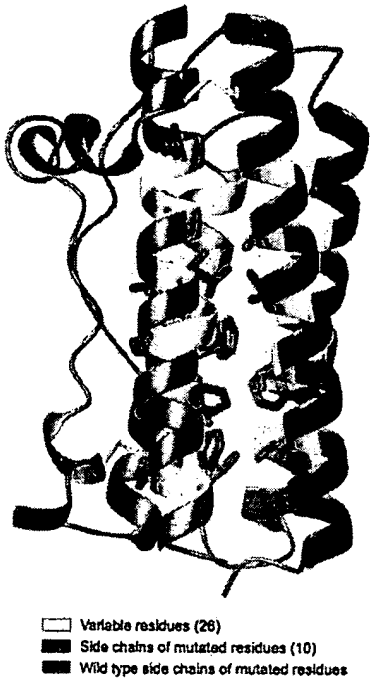

Fig. 3. Modeled structure of hG-CSF analog (core10) obtained from deep core design. Twenty-six core residues were allowed to vary; computational screening with PDA resulted in 10 mutations: C17L, G28A, L78F, Y85F, L103V, V110I, F113L, V151I, V153I, and L168F.

To determine the importance of the other mutations, another sequence was made (core2) that contained only two of the core10 mutations, G28A and C17A; all other residues were identical to wild type (Fig. 2). The T_m of core2 was 5°C higher than wild type, indicating that improvements in helical propensity and the elimination of a free cysteine are important for heightened thermostability. The remainder of the increase in T_m seen for core10 may be attributable to improved packing interactions and increased hydrophobic burial.

Storage stability

Increased shelf life is important for distribution and storage and is a desirable feature for G-CSF and other protein drugs. Because aggregation and chemical degradation are the predominant mechanisms of inactivation of G-CSF (Herman et al. 1996), shelf life was estimated by incubating the proteins at elevated temperature and then using size-exclusion chromatography to observe the disappearance of monomeric protein. Chemical degradation was estimated using reverse phase chromatography (data not shown). Core2 and core10 showed five- and 10-fold improvements in storage stability, respectively, at 50°C (Fig. 5). Rate constants were determined by a first order exponential fit of the fraction monomer remaining/time curves using KaleidaGraph (Synergy Software).

Biological activity

Granulopoietic activity was determined in vitro by quantitating cell proliferation as a function of protein concentration in murine lymphoid cells transfected with the gene for the human G-CSF receptor. The designed proteins were as

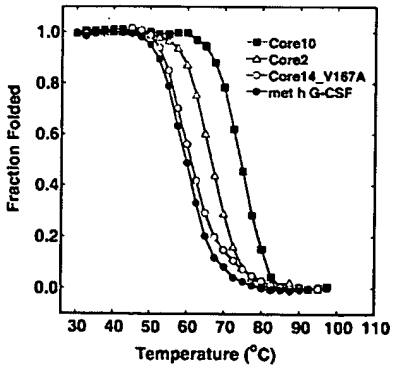

Fig. 4. Thermal stability of hG-CSF analogs. Thermal stability was assessed by monitoring the temperature dependence of the circular dichroism spectral signal at 222 nm. Melting temperatures (T_ms) were derived from the derivative curve of the ellipticity at 222 nm versus temperature. Core10 and core2 showed increases in T_m of 13°C and 5°C, respectively, over native met hG-CSF.

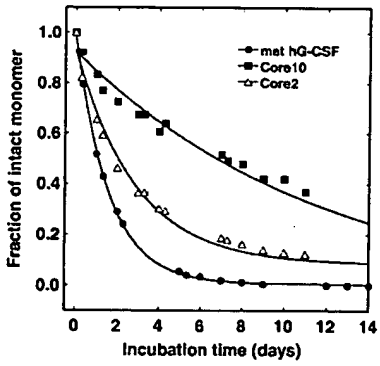

Fig. 5. Shelf life of hG-CSF analogs. Shelf life was estimated by incubating the proteins at elevated temperature (50°C) and using size exclusion chromatography to observe disappearance of monomeric protein. Rate constants were determined by a first order exponential fit of the fraction monomer remaining/time curves. Core2 and core10 showed five- and 10-fold improvements in storage stability, respectively, over met hG-CSF controls.

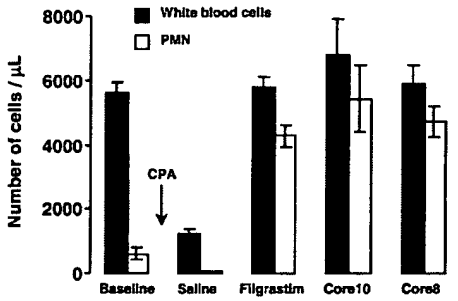

Fig. 6. In vivo granulopoietic activity of hG-CSF analogs. Mice were rendered neutropenic with a single intraperitoneal injection of 200 mg/kg cyclophosphamide (CPA). Beginning 24 h later and for 4 consecutive days, the mice were given a daily intravenous injection of 100 µg/kg of native hG-CSF (filgrastim, Amgen), an hG-CSF analog, or saline. On day 5, granulopoietic activity was determined by counting the number of white blood cells and polymorphonuclear neutrophils (PMN). The designed analogs (core8 and core10) were as effective as controls in eliciting a granulopoietic response.

active as wild-type hG-CSF (data not shown). The designed analogs were also as effective as wild type in increasing white blood cell and polymorphonuclear neutrophil levels in the neutropenic mouse (Fig. 6). Neutropenia, characterized by an abnormally low level of neutrophils in the blood, was induced by injection of cyclophosphamide. Reversal of this effect by the designed analogs shows that granulopoietic activity was also retained in vivo.

Pharmacokinetics

The pharmacokinetics of core10 and native hG-CSF (filgrastim, Amgen) was studied in cynomolgus monkeys after a single subcutaneous or intravenous injection of 5 µg/kg and after daily subcutaneous injections of 5 µg/kg for 28 d. Analysis of the serum concentration-time curves shows that subcutaneous injection of the designed analog results in greater systemic exposure (area under concentration-time curve, AUC) than the same dose of wild-type hG-CSF (Fig. 7B). This was true after a single dose on day 1 (78.8 vs. 54.6 ng-h/mL, data not shown), as well as on day 28 (37.2 vs. 17.4 ng-h/mL). There were no measurable differences in serum half-life. In the intravenous study, however, the half-life of core10 was three-fold shorter (1 vs. 3 h), and the AUC was significantly less (54.7 vs. 117.4 ng-h/mL), indicating that core10 is cleared faster (Fig. 7A). Taken together, these data indicate that the designed analog is absorbed more quickly from the subcutaneous compartment (absorption could not be measured directly given the small number of data points at early times). Improved absorption may be attributable to decreased aggregation or association of the designed protein. The increased monomer lifetime and decreased aggregation seen in our shelf-life studies and the improved thermal stability of the native conformation observed for core10 indicate a decrease in aggregation in the subcutaneous compartment. This possibility is supported by the fact that other protein therapeutics engineered for reduced aggregation also show faster absorption rates. For example, insulin Lispro and other rapid-acting insulin analogs that were designed to decrease their tendency to self-associate are absorbed faster than regular insulin after subcutaneous injection (Howey et al. 1994; Home et al. 1999).

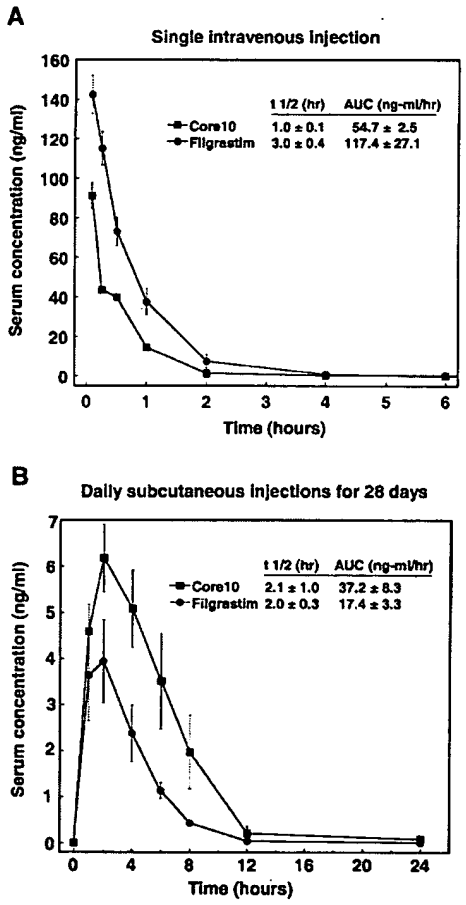

Fig. 7. Pharmacokinetics of hG-CSF analogs. Plasma concentrations of a designed hG-CSF analog or wild-type hG-CSF (filgrastim, Amgen) were determined after administration in cynomolgus monkeys. (*A*) Animals were given a single intravenous injection of 5 µg/kg or (*B*) daily subcutaneous injections of 5 µg/kg for 28 d. Noncompartmental analysis of the serum concentration-time curves shows that subcutaneous injections of the core10 analog resulted in greater systemic exposure (area under concentration-time curve, AUC) than the same dose of wild-type hG-CSF, whereas there was no change in serum half-life ($t_{1/2}$). In the intravenous study, the AUC was significantly less and the $t_{1/2}$ three-fold shorter, indicating that core10 was cleared faster.

Comparison to published G-CSF variants

In vitro and cassette mutagenesis studies have shown that alterations of the N-terminal region of G-CSF can lead to improved granulopoietic activity (Kuga et al. 1989; Okabe et al. 1990). Point mutations at Cys17 have also been found to affect shelf life; replacement with Ala led to an increase, Ser had no effect, and large residues (Ile, Tyr, Arg) led to a decrease (Ishikawa et al. 1992). In contrast, our core10 sequence, which has a large residue (Leu) at this position, showed an improved shelf life. This may be explained by the observation that in a Cys17Leu point mutant, Leu's side chain would clash with the aromatic ring of the nearby Phe at position 113. This steric clash does not occur in core10, however, because the Phe at 113 is replaced by Leu and, in compensation for this change, two nearby Leu's become Phe's (at positions 78 and 168). Thus, multiple mutations allow complementary repacking of the hydrophobic core in the core10 mutant and may be responsible for its enhanced stability and shelf life.

Significant improvements in thermal stability were also observed when the seven helical Gly residues in G-CSF were replaced with Ala to form point, double, and triple mutants (Bishop et al. 2001). Substitutions at positions 26, 28, 149, and 150 were the most effective. The investigators attributed the stabilizing effect to the enhancement in α-helical propensity associated with the Gly/Ala substitutions. These data support our suggestion that the heightened thermal stability seen with our mutants (which also contain a Gly/Ala substitution at position 28) is at least in part attributable to an improvement in helical propensity.

Probing the robustness of PDA with a homology modeled core position

As pointed out previously, the homology modeling of human G-CSF onto the bovine structure was straightforward for the most part because the replaced residues were primarily solvent exposed and no rearrangement of the backbone was necessary. The change at one core position, however, G167V, induced a steric clash and energy minimization of the entire protein was used to relieve the strain. We decided to assess the impact of this manipulation by doing an additional design (core167V) in which the variable residues were essentially the same as in the deep core design except that position 167 was also allowed to vary. We found that Val167 mutated to Ala (the other mutations were essentially the same as for core10). To probe the plasticity of the core, instead of using this PDA optimal sequence, which only had two mutations in this region, we ran experiments on another high-scoring sequence (core14_V167A) that had additional mutations (14 total, including L157I, F160W, and L161F). This sequence was chosen because it balanced an extensive number of mutations with a relatively high design score. Although it ranked 21st in the sequence energy list and was 2 kcal/mole less favorable than the optimal sequence, it was still biologically active and as stable as wild type (T_m of 61°C) (Figs. 2, 4). This indicates that optimization with PDA is fairly robust, and that the protein core can be quite plastic and can accommodate large changes without sacrificing stability or function.

Conclusions

PDA is a powerful ultrahigh throughput computational screening method. Its ability to screen up to 10^{80} sequences and allow multiple simultaneous mutations significantly increases the likelihood of finding new and improved proteins. In this study, PDA was used to develop improved analogs for a therapeutically important protein, hG-CSF. The novel proteins showed enhanced thermal stabilities and shelf life while retaining biological activity. Analysis of the mutants and results obtained with derived sequences indicates that the heightened stability is attributable to improvements in helical propensity and the elimination of a free cysteine; improved core packing and optimized hydrophobic burial of side chains may also be important. Pharmacokinetic studies indicate that subcutaneous injection of the most stable variant results in greater systemic exposure, probably attributable to improved absorption from the subcutaneous compartment.

These results show that PDA can be successfully applied to proteins of therapeutic interest. They also illustrate the value of its precise control over the site and type of mutations, allowing for the rational design of desired properties such as improved stability and pharmacokinetics and the elimination of undesirable ones such as toxicity and antigenicity. These features are particularly important in the design of therapeutic proteins. PDA thus has great potential as a powerful in silico tool for therapeutic protein design.

Materials and methods

Template structure preparation

The template structure for the designed proteins was produced by homology modeling using the crystal structure of bovine G-CSF (Brookhaven Protein Data Bank code 1bgc) as the starting point. The program BIOGRAF (Molecular Simulations Inc., San Diego, CA) was used to generate explicit hydrogens on the structure, which was then minimized for 50 steps using the conjugate gradient method and the Dreiding II force field (Mayo et al. 1990). The residues that differ in the bovine sequence or were not present in the bovine crystal structure were replaced with the human residues for those positions. The conformations of the replaced side chains were optimized using PDA (Dahiyat and Mayo 1997a,b), and the entire structure was minimized again for 50 steps. This minimized structure was used as the template for all the designs.

Protein design

Analogs of hG-CSF were designed by simultaneously optimizing residues in the buried core of the protein using PDA. The computational details, residue classification, potential functions, and parameters used for van der Waals interactions, solvation, and hydrogen bonding are described in previous work (Dahiyat and Mayo 1996, 1997a). An expanded version of the backbone-dependent rotamer library of Dunbrack and Karplus (Dunbrack and Karplus 1993) was used in all the calculations. The global optimum sequence from each design was selected for characterization and experimental testing, except for core167V in which the 21st ranked sequence was used. Calculations were generally performed overnight using 16 processors of an SGI Origin 2000 with 32 R10000 processors running at 195 MHz. The length of the runs varied from 1 to several hours of CPU time.

Cloning and expression

A gene for met hG-CSF was synthesized from partially overlapping oligonucleotides (~100 bases) that were extended and PCR amplified. Codon usage was optimized for *E. coli* and several restriction sites were incorporated to ease future cloning. These partial genes were cloned into a vector and transformed into *E. coli* for sequencing. Several of these gene fragments were then cloned into adjacent positions in an expression vector (pET17 or pET21) to form the full-length gene for met hG-CSF (528 bases) and transformed into *E. coli* for expression. Protein was expressed in *E. coli* in insoluble inclusion bodies and its identity was confirmed by immunoblot of SDS-PAGE using a commercial mAb against hG-CSF.

Refolding, purification, and storage

The protein inclusion bodies were solubilized in detergent and refolded in the presence of $CuSO_4$ to promote formation of native disulfide bonds (Lu et al. 1992). A size-exclusion column (10 mm × 300 mm loaded with Superdex prep 75 resin purchased from Pharmacia) was loaded with protein and eluted at a flow rate of 0.8 mL/min using the column buffer (100 mM Na_2SO_4, 50 mM Tris, pH 7.5). The peaks were monitored at dual wavelengths of 214 nm and 280 nm. Albumin, carbonic anhydrate, cytochrome C, and aprotinin were used to calibrate the molecular size of proteins versus elution time. The monomeric peak that elutes around the expected elution time for each protein was collected and the buffer was exchanged into 10 mM NaOAc at pH 4 for biophysical characterization. For long-term storage, a buffer of 5% sorbitol, 0.004% Tween 80, and 10 mM NaOAc at pH 4 was used. A pH of 4 was chosen for these buffers to be consistent with the commercial formulation of hG-CSF (Amgen), which was used as a control. The proteins were >98% pure as judged by reversed phase high performance liquid chromatography (HPLC) on a C4 column (3.9 mm × 150 mm) with a linear acetonitrile-water gradient containing 0.1% TFE. The identities of all proteins were confirmed by comparing the molecular mass measured by mass spectrometry with corresponding molecular mass calculated using the protein sequences.

Spectroscopic characterization

Protein samples were 50 μM in 50 mM sodium phosphate at pH 5.5. Concentrations were determined using UV spectrophotometry. Protein structure was assessed by CD. CD spectra were measured on an Aviv 202DS spectrometer equipped with a Peltier temperature control unit using a 1-mm path length cell. Thermal stability was assessed by monitoring the temperature dependence of the CD signal at 222 nm (Kolvenbach et al. 1997). A buffer of 10 mM NaOAc was used at pH 4.0 and data were collected every 2.5°C with an averaging time of 5 sec and an equilibration time of 3 min. Thermal denaturation curves were smoothed using KaleidaGraph. The melting temperature (T_m) of each protein was derived from the derivative curve of the ellipticity at 222 nm versus temperature. The T_m values were reproducible to within 2°C for the same protein at the concentrations used.

Storage stability

The storage stability of the designed proteins was assessed by incubation at both 37°C and 50°C under solution conditions identical to that used in the commercial formulation of hG-CSF (filgrastim, Amgen). Because aggregation and chemical degradation are the predominant mechanisms of inactivation of G-CSF (Herman et al. 1996), accelerated degradation was followed by observing the disappearance of monomeric protein with both size-exclusion and reverse-phase chromatography. Rate constants for shelf-life estimation were determined by a first-order exponential fit of the fraction monomer remaining/time curves using KaleidaGraph (Synergy Software).

Cell proliferation assay

Granulopoietic activity was measured by quantifying cell proliferation as a function of protein concentration using Ba/F3 (murine lymphoid) cells stably transfected with the gene encoding the human Class 1 G-CSF receptor (Avalos et al. 1995). Cell proliferation was detected by 5-bromo-2′-deoxyuridine (BrdU) incorporation quantified by a BrdU-specific ELISA kit (Boehringer Mannheim).

In vivo biological activity

Granulopoietic activity was determined in the neutropenic mouse (Hattori et al. 1990). C57BL/6 mice were rendered neutropenic with a single intraperitoneal injection of 200 mg/kg cyclophosphamide (CPA). Beginning 24 h later and for 4 consecutive days, the mice were given a daily intravenous injection of 100 μg/kg of an hG-CSF analog, met hG-CSF produced in our laboratory, clinically available hG-CSF (filgrastim, Amgen), or saline. On day 5, 6 h after the final dose, the animals were killed, blood samples were collected, and granulopoietic activity was determined by counting the number of white blood cells and polymorphonuclear neutrophils.

Pharmacokinetics

Plasma concentrations of a designed hG-CSF analog or wild-type hG-CSF (filgrastim, Amgen) were determined following administration in cynomolgus monkeys. Animals were given a single intravenous injection of 5 μg/kg or daily subcutaneous injections of 5 μg/kg for 28 d. In the intravenous study, blood samples were collected at 0 (predose), 5, 15, and 30 min and 1, 2, 4, 6, 8, 12, and 24 h postdosing. In the subcutaneous studies, blood samples were collected at 0 (predose), 1, 2, 4, 6, 8, 12, and 24 h postdosing on day 1 and day 28. All samples were immediately placed on wet ice and centrifuged at 28°C. The resultant plasma was then frozen and

stored (−70°C). Plasma concentrations were determined using an enzyme-linked immunosorbent assay (Quantikine human G-CSF ELISA, R&D Systems, Minneapolis, MN), performed per manufacturers instructions except that samples were diluted in PBS, 5% nonfat dry milk, and 0.05% Tween 20, and the incubation was extended to overnight at 4°C. Plasma concentrations of the designed hG-CSF analog and filgrastim were estimated from their corresponding standard curves. Pharmacokinetic parameters were calculated by noncompartmental analysis. The terminal slope (λz) was estimated by linear regression through the last time points of the log concentration versus time curves and used to calculate the terminal half-life ($t_{1/2}$). The area under the curve from time of dosing through the last time point (AUC_{0-z}) was calculated by the linear trapezoid method.

Acknowledgments

We thank Dr. Belinda Avalos (Ohio State University) for kindly supplying the Ba/F3 cell line transfected with the hG-CSF receptor. We also thank Dr. Steven Adams (American College of Laboratory Animal Medicine) and LAB Preclinical Research Institute Inc., (Quebec, Canada) for conducting the monkey studies.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Aritomi, M., Kunishima, N., Okamoto, T., Kuroki, R., Ota, Y., and Morikawa, K. 1999. Atomic structure of the GCSF-receptor complex showing a new cytokine-receptor recognition scheme. *Nature* **401:** 713–717.

Avalos, B.R., Hunter, M.G., Parker, J.M., Ceselski, S.K., Druker, B.J., Corey, S.J., and Mehta, V.B. 1995. Point mutations in the conserved box 1 region inactivate the human granulocyte colony-stimulating factor receptor for growth signal transduction and tyrosine phosphorylation of p75c-rel. *Blood* **85:** 3117–3126.

Bishop, B., Koay, D.C., Sartorelli, A.C., and Regan, L. 2001. Reengineering granulocyte colony-stimulating factor for enhanced stability. *J. Biol. Chem.* **276:** 33465–33470.

Bolon, D.N. and Mayo, S.L. 2001. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci.* **98:** 14274–14279.

Chang, C.C., Chen, T.T., Cox, B.W., Dawes, G.N., Stemmer, W.P., Punnonen, J., and Patten, P.A. 1999. Evolution of a cytokine using DNA family shuffling. *Nat. Biotechnol.* **17:** 793–797.

Chen, K.Q. and Arnold, F.H. 1991. Enzyme engineering for nonaqueous solvents: Random mutagenesis to enhance activity of subtilisin E in polar organic media. *Biotechnology* **9:** 1073–1077.

Crameri, A., Whitehorn, E.A., Tate, E., and Stemmer, W.P. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14:** 315–319.

Dahiyat, B.I. and Mayo, S.L. 1996. Protein design automation. *Protein Sci.* **5:** 895–903.

———. 1997a. De novo protein design: Fully automated sequence selection. *Science* **278:** 82–87.

———. 1997b. Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci.* **94:** 10172–10177.

Desjarlais, J.R. and Handel, T.M. 1995. De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4:** 2006–2018.

Dunbrack, R.L. and Karplus, M. 1993. Backbone-dependent rotamer library for proteins—an application to side-chain prediction. *J. Mol. Biol.* **230:** 543–574.

Grossmann, M., Leitolf, H., Weintraub, B.D., and Szkudlinski, M.W. 1998. A rational design strategy for protein hormone superagonists. *Nat. Biotechnol.* **16:** 871–875.

Hattori, K., Shimizu, K., Takahashi, M., Tamura, M., Oheda, M., Ohsawa, N., and Ono, M. 1990. Quantitative in vivo assay of human granulocyte colony-stimulating factor using cyclophosphamide-induced neutropenic mice. *Blood* **75:** 1228–1233.

Heikoop, J.C., van den Boogaart, P., Mulders, J.W., and Grootenhuis, P.D. 1997. Structure-based design and protein engineering of intersubunit disulfide bonds in gonadotropins. *Nat. Biotechnol.* **15:** 658–662.

Hellinga, H.W. and Richards, F.M. 1994. Optimal sequence selection in proteins of known structure by simulated evolution. *Proc. Natl. Acad. Sci.* **91:** 5803–5807.

Herman, A.C., Boone, T.C., and Lu, H.S. 1996. Characterization, formulation, and stability of Neupogen (Filgrastim), a recombinant human granulocyte-colony stimulating factor. *Pharm. Biotechnol.* **9:** 303–328.

Hill, C.P., Osslund, T.D., and Eisenberg, D. 1993. The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors. *Proc. Natl. Acad. Sci.* **90:** 5167–5171.

Home, P.D., Barriocanal, L., and Lindholm, A. 1999. Comparative pharmacokinetics and pharmacodynamics of the novel rapid-acting insulin analogue, insulin aspart, in healthy volunteers. *Eur. J. Clin. Pharmacol.* **55:** 199–203.

Horan, T., Wen, J., Narhi, L., Parker, V., Garcia, A., Arakawa, T., and Philo, J. 1996. Dimerization of the extracellular domain of granuloycte-colony stimulating factor receptor by ligand binding: A monovalent ligand induces 2:2 complexes. *Biochemistry* **35:** 4886–4896.

Howey, D.C., Bowsher, R.R., Brunelle, R.L., and Woodworth, J.R. 1994. [Lys(B28), Pro(B29)]-human insulin. A rapidly absorbed analogue of human insulin. *Diabetes* **43:** 396–402.

Ishikawa, M., Iijima, H., Satake-Ishikawa, R., Tsumura, H., Iwamatsu, A., Kadoya, T., Shimada, Y., Fukamachi, H., Kobayashi, K., Matsuki, S., et al. 1992. The substitution of cysteine 17 of recombinant human G-CSF with alanine greatly enhanced its stability. *Cell Struct. Funct.* **17:** 61–65.

Jiang, X., Farid, H., Pistor, E., and Farid, R.S. 2000. A new approach to the design of uniquely folded thermally stable proteins. *Protein Sci.* **9:** 403–416.

Kolvenbach, C.G., Narhi, L.O., Philo, J.S., Li, T., Zhang, M., and Arakawa, T. 1997. Granulocyte-colony stimulating factor maintains a thermally stable, compact, partially folded structure at pH2. *J. Pept. Res.* **50:** 310–318.

Kraemer-Pecore, C.M., Wollacott, A.M., and Desjarlais, J.R. 2001. Computational protein design. *Curr. Opin. Chem. Biol.* **5:** 690–695.

Kuga, T., Komatsu, Y., Yamasaki, M., Sekine, S., Miyaji, H., Nishi, T., Sato, M., Yokoo, Y., Asano, M., Okabe, M., et al. 1989. Mutagenesis of human granulocyte colony stimulating factor. *Biochem. Biophys. Res. Commun.* **159:** 103–111.

Lovejoy, B., Cascio, D., and Eisenberg, D. 1993. Crystal structure of canine and bovine granulocyte-colony stimulating factor (G-CSF). *J. Mol. Biol.* **234:** 640–653.

Lowman, H.B. and Wells, J.A. 1993. Affinity maturation of human growth hormone by monovalent phage display. *J. Mol. Biol.* **234:** 564–578.

Lu, H.S., Clogston, C.L., Narhi, L.O., Merewether, L.A., Pearl, W.R., and Boone, T.C. 1992. Folding and oxidation of recombinant human granulocyte colony stimulating factor produced in *Escherichia coli*. Characterization of the disulfide-reduced intermediates and cysteine—serine analogs. *J. Biol. Chem.* **267:** 8770–8777.

Malakauskas, S. and Mayo, S. 1998. Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* **5:** 470–475.

Marshall, S.A. and Mayo, S.L. 2001. Achieving stability and conformational specificity in designed proteins via binary patterning. *J. Mol. Biol.* **305:** 619–631.

Mayo, S.L., Olafson, B.D., and Goddard III, W.A. 1990. Dreiding: A generic forcefield for molecular simulations. *J. Phys. Chem.* **94:** 8897–8909.

Okabe, M., Asano, M., Kuga, T., Komatsu, Y., Yamasaki, M., Yokoo, Y., Itoh, S., Morimoto, M., and Oka, T. 1990. In vitro and in vivo hematopoietic effect of mutant human granulocyte colony-stimulating factor. *Blood* **75:** 1788–1793.

Pokala, N. and Handel, T.M. 2001. Review: Protein design—where we were, where we are, where we're going. *J. Struct. Biol.* **134:** 269–281.

Reidhaar-Olson, J.F., De Souza-Hart, J.A., and Selick, H.E. 1996. Identification of residues critical to the activity of human granulocyte colony-stimulating factor. *Biochemistry* **35:** 9034–9041.

Shimaoka, M., Shifman, J.M., Jing, H., Takagi, J., Mayo, S.L., and Springer, T.A. 2000. Computational design of an integrin I domain stabilized in the open high affinity conformation. *Nat. Struct. Biol.* **7:** 674–678.

Stemmer, W.P. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370:** 389–391.

Street, A.G. and Mayo, S.L. 1999. Computational protein design. *Structure Fold. Des.* **7:** R105–109.

Strop, P. and Mayo, S.L. 1999. Rubredoxin variant folds without iron. *J. Am. Chem. Soc.* **121:** 2341–2345.

Young, D.C., Zhan, H., Cheng, Q.L., Hou, J., and Matthews, D.J. 1997. Characterization of the receptor binding determinants of granulocyte colony stimulating factor. *Protein Sci.* **6:** 1228–1236.

Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16:** 258–261.

47. (New) A molecular library comprising a plurality of members, each member comprising a recombinant nucleic acid, wherein each of said members comprises a fusion nucleic acid comprising, from 5' to 3':

i) a first nucleic acid encoding a first dimerization peptide;
ii) a second nucleic acid encoding a random peptide; and
iii) a third nucleic acid encoding a second dimerization peptide;
wherein each of said random peptides is different.

48. (New) A molecular library according to claim 47, wherein at least one said dimerization peptide is FLIVK.

49. (New) A molecular library according to claim 47, wherein at least one said dimerization peptide is KFLIVKS.

50. (New) A molecular library according to claim 47, wherein at least one said dimerization peptide is FLIVE.

51. (New) A molecular library according to claim 47, wherein said first dimerization peptide is FLIVK and said second dimerization peptide is FLIVE.

52. (New) A cellular library comprising a plurality of cells, each cell comprising a member of the molecular library of claim 47.--